

# DEVICE AND METHOD FOR MODEL ADAPTATION, RECORDING MEDIUM, AND VOICE RECOGNITION DEVICE

**Publication number:** JP2002156992 (A)

**Publication date:** 2002-05-31

**Inventor(s):** NAKATSUKA KOUCHIYO +

**Applicant(s):** SONY CORP +

**Classification:**

- international: **G10L15/06; G10L15/20; G10L15/00; (IPC1-7): G10L15/06; G10L15/20**

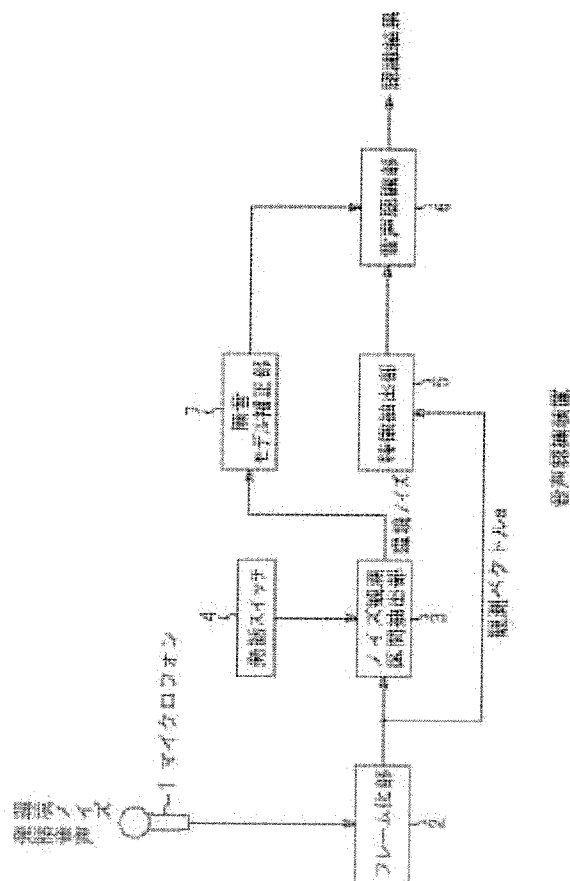
- **European:**

**Application number:** JP20000353790 20001121

**Priority number(s):** JP20000353790 20001121

Abstract of JP 2002156992 (A)

**PROBLEM TO BE SOLVED:** To adapt a soundless model which is sufficiently adaptive to a soundless part in a voice section. **SOLUTION:** A noise observation section extraction part 3 extracts environment noise observed in the section right before a sound section and supplies it to a soundless model correction part 7. The soundless model correction part 7 adapts a soundless model as a sound model showing no sound according to the environment noise in the section right before the sound section of a current voice to be recognized and environment noise in the section right before a sound section of a voice as an object of past voice recognition.



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号  
特開2002-156992  
(P2002-156992A)

(43)公開日 平成14年 5月31日 (2002.5.31)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テ-マ-ト* (参考)
G 1 0 L 15/06		G 1 0 L 3/00	5 2 1 T 5 D 0 1 5
15/20			5 3 1 Q

審査請求 未請求 請求項の数11 O L (全 23 頁)

(21)出願番号 特願2000-353790(P2000-353790)

(22)出願日 平成12年11月21日(2000. 11. 21)

(71)出願人 000002185

ソニー株式会社

東京都品川区北品川 6 丁目 7 番35号

(72)発明者 中塚 洪長

東京都品川区北品川 6 丁目 7 番35号 ソニ  
ー株式会社内

(74)代理人 100082131

弁理士 稲本 義雄

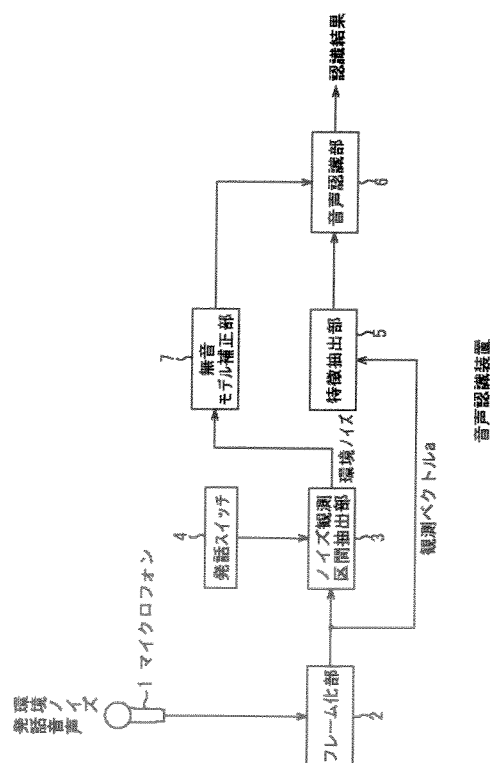
Fターム(参考) 5D015 EE05 GG00

(54)【発明の名称】 モデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置

(57)【要約】

【課題】 音声区間中の無音部分に十分対処可能な無音モデルの適応を行う。

【解決手段】 ノイズ観測区間抽出部3は、音声区間の直前の区間で観測される環境ノイズを抽出し、無音モデル補正部7に供給する。無音モデル補正部7は、現在の音声認識の対象となっている音声の音声区間の直前の区間における環境ノイズと、過去に音声認識の対象とされた音声の音声区間の直前の区間における環境ノイズとに基づいて、無音を表す音響モデルである無音モデルの適応を行う。



## 【特許請求の範囲】

【請求項1】 音声を認識するのに用いる音響モデルの適応を行うモデル適応装置であって、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とするモデル適応装置。

【請求項2】 前記モデル適応手段は、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた1以上の音声の音声区間の直前の区間における抽出データとから、現在の音声認識の対象となっている音声の認識に用いる前記無音モデルを生成することを特徴とする請求項1に記載のモデル適応装置。

【請求項3】 前記モデル適応手段は、過去に音声認識の対象とされた1以上の音声の音声区間の直前の区間における抽出データに基づいて、第1の無音モデルを生成するとともに、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データに基づいて、第2の無音モデルを生成し、前記第1と第2の無音モデルに基づいて、現在の音声認識の対象となっている音声の認識に用いる前記無音モデルを生成することを特徴とする請求項1に記載のモデル適応装置。

【請求項4】 前記モデル適応手段は、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データに基づいて、前記無音モデルを生成し、その無音モデルと、過去に音声認識の対象とされた音声の認識に用いられた前記無音モデルとに基づいて、現在の音声認識の対象となっている音声の認識に用いる前記無音モデルを生成することを特徴とする請求項1に記載のモデル適応装置。

【請求項5】 前記音声の認識は、音声の特徴空間における特徴量のベクトルまたは特徴量の分布に基づいて行われ、前記モデル適応手段は、前記抽出データから得られる前記特徴量のベクトルまたは特徴量の分布に基づいて、前記無音モデルの適応を行うことを特徴とする請求項1に記載のモデル適応装置。

【請求項6】 前記モデル適応手段は、前記抽出データから得られる前記特徴量と特徴量の分布の両方に基づいて、前記無音モデルの適応を行うことを特徴とする請求項5に記載のモデル適応装置。

【請求項7】 前記モデル適応手段は、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データそれぞれに重

みを付して、前記無音モデルの適応を行うことを特徴とする請求項1に記載のモデル適応装置。

【請求項8】 前記モデル適応手段は、統計的手法によって、前記無音モデルの適応を行うことを特徴とする請求項1に記載のモデル適応装置。

【請求項9】 音声を認識するのに用いる音響モデルの適応を行うモデル適応方法であって、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えることを特徴とするモデル適応方法。

【請求項10】 音声を認識するのに用いる音響モデルの適応を行うモデル適応処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

【請求項11】 音声を、音響モデルを用いて認識する音声認識装置であって、音声データの特徴量を抽出する特徴抽出手段と、前記特徴量と音響モデルに基づいて、前記音声を認識する音声認識手段と、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とする音声認識装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、モデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置に関し、特に、例えば、ノイズに起因する音声認識性能の劣化を防止することができるようにするモデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置に関する。

## 【0002】

【従来の技術】 音声認識装置においては、例えば、音声認識対象の音声から、その特徴ベクトルが抽出され、そ

10

20

30

40

50

の特徴ベクトルの系列が、音声の音響モデルから観測される尤度を計算すること等によって、音声認識される。

#### 【0003】

【発明が解決しようとする課題】ところで、音声認識装置においては、一般に、ユーザの発話が行われている区間である音声区間を特定し、その音声区間を対象に、音声認識が行われる。

【0004】しかしながら、ユーザの音声は、音声区間の全体にわたって存在するとは限らない。即ち、音声区間には、一般に、息継ぎ等によって、ユーザの音声が存在しない部分がある。

【0005】一方、音声認識装置が使用される環境においては、認識対象の音声以外の音、即ち、ノイズが存在する。

【0006】具体的には、例えば、音声を入力するマイク（マイクロフォン）を叩く音や、場所によっては、ドアを開閉する音、ユーザの咳の音、音声認識しようとしている音声のユーザ以外のユーザの発話等が、ノイズとして存在する。また、例えば、音声認識装置が、エンターテイメント用のロボット等に適用された場合には、そのロボットに各種の動作を行わせるためのアクチュエータの音が、ノイズとして存在し、さらに、そのロボットが、デモンストレーション会場で公表されるときには、観衆の話し声や拍手等が、ノイズとして存在する。

【0007】従って、音声区間において、ユーザの音声が存在しない部分には、上述したようなノイズのみが存在することとなるが、音声認識装置では、そのノイズのみの部分についても、ユーザの音声が存在するものとして、音響モデルを用いて、音声認識が行われるため、認識性能が劣化することがあった。即ち、特に、音声区間の開始から、実際に、ユーザの発話が始まるまでの時間が長くなると、認識性能が低下する課題があった。

【0008】そこで、ユーザの音声が存在しない状態、即ち、音声認識装置が使用される環境においてノイズが存在する場合には、そのノイズのみが存在する状態としての無音を表す音響モデルである無音モデルを導入し、音声区間の中で、ユーザの音声が存在しない部分（以下、適宜、無音部分という）については、その無音モデルで対処する方法がある。

【0009】しかしながら、音声認識装置が使用される環境におけるノイズは、一定であるとは限らず、むしろ時々刻々と変化することが多いため、あらかじめ作成しておいた無音モデルを、そのまま用いるのでは、音声区間中の無音部分について、十分に対処することができない場合がある。

【0010】そこで、本件出願人は、例えば、特開2000-259198号公報（特願平11-57467号）において、音声区間の直前の区間における音声（ノイズ）に基づいて、無音モデルの適応を行う方法について、先に提案してい

る。

【0011】しかしながら、先に提案した方法では、現在の音声認識の対象となっている音声の音声区間（以下、適宜、注目音声区間という）の直前の区間における音声にのみ基づいて、無音モデルの適応を行うため、例えば、注目音声区間の直前において、ユーザが、音声の入力に用いるマイクを叩く等した場合や、観衆が拍手を行った場合等の、いわば突発的なノイズが生じた場合、その突発的なノイズに基づいて、無音モデルの適応が行われることがあり、この場合、音声区間中の無音部分について、十分に対処することが困難であると考えられる。

【0012】また、そのような突発的なノイズが生じず、比較的定常的なノイズが長時間連続している場合には、注目音声区間のみならず、過去に音声認識の対象とされた音声の音声区間の直前の区間におけるノイズをも用いて、無音モデルの適応を行った方が、音声区間中の無音部分について、より十分に対処することができると予想される。

【0013】本発明は、このような状況に鑑みてなされたものであり、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことができるようにし、これにより、無音部分に起因する音声認識性能の劣化を防止（低減）することができるようにするものである。

#### 【0014】

【課題を解決するための手段】本発明のモデル適応装置は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とする。

【0015】本発明のモデル適応方法は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えることを特徴とする。

【0016】本発明の記録媒体は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えるプログラムが記録されているこ

とを特徴とする。

【0017】本発明の音声認識装置は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とする。

【0018】本発明のモデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置においては、音声区間の直前の区間で観測される音声データが抽出され、抽出データとして出力される。そして、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応が行われる。

【0019】

【発明の実施の形態】図1は、本発明を適用した音声認識装置の一実施の形態の構成例を示している。

【0020】この音声認識装置において、マイク1は、認識対象である発話音声と、環境ノイズとともに集音し、フレーム化部2に出力する。フレーム化部2は、マイク1から入力される音声データを、所定の時間間隔（例えば、10ms）で取り出し、その取り出したデータを、1フレームのデータとして出力する。フレーム化部2が出力する1フレーム単位の音声データは、そのフレームを構成する時系列の音声データそれぞれをコンポーネントとする観測ベクトル $a$ として、ノイズ観測区間抽出部3、および特徴抽出部5に供給される。

【0021】ここで、以下、適宜、第 $t$ フレームの音声データである観測ベクトルを、 $a(t)$ と表す。

【0022】ノイズ観測区間抽出部3は、フレーム化部2から入力されるフレーム単位の音声データを所定の時間（ $M$ フレーム分以上）だけバッファリングし、図2に示すように、発話スイッチ4がオンとされるタイミング $t_0$ から $M$ フレーム分だけ以前のタイミング $t_m$ までをノイズ観測区間 $T_n$ として、そのノイズ観測区間 $T_n$ における $M$ フレーム分の観測ベクトル $a$ を抽出して、無音モデル補正部7に出力する。

【0023】発話スイッチ4は、ユーザが発話を開始するときにユーザによってオンとされ、発話を終了するときにオフとされる。したがって、発話スイッチ4がオンとされたタイミング $t_0$ 以前（ノイズ観測区間 $T_n$ ）の音声データには、発話音声は含まれず、環境ノイズだけが存在する。また、発話スイッチ4がオンとされたタイミング $t_0$ から発話スイッチ4がオフとされるタイミング $t_m$ までは、音声区間とされて、その音声区間の音声データが音声認識の対象とされる。

【0024】特徴抽出部5は、例えば、図3に示すように、パワースペクトラム分析部11から構成され、フレーム化部2からの音声区間における観測ベクトル $a$ としての音声データをフーリエ変換することにより、音声の特徴量として、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトル $y$ を算出する。特徴抽出部5で得られた特徴ベクトル $y$ は、音声認識部6に供給される。

【0025】なお、パワースペクトラムの算出方法は、フーリエ変換によるものに限定されるものではない。すなわち、パワースペクトラムは、その他、例えば、いわゆるフィルタバンク法などによって求めることも可能である。

【0026】また、ここでは、音声の特徴量として、パワースペクトラムを用いることとしているが、音声の特徴量としては、パワースペクトラムの他、ケプストラム係数（MFCC（Mel Frequency Cepstrum Coefficient）を含む）や、線形予測係数その他を採用することが可能である。

【0027】音声認識部6は、特徴抽出部5から入力される特徴ベクトル $y$ を、所定数 $K$ の音響モデルと1個の無音モデルの中のいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。すなわち、音声認識部6は、例えば、無音区間に対応する識別関数（特徴パラメータ $y$ が無音モデルに分類されるかを識別するための関数）と、所定数 $K$ の単語それぞれに対応する識別関数（特徴パラメータ $y$ がいずれの音響モデルに分類されるかを識別するための関数）とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部5からの特徴ベクトル $y$ を引数として計算する。そして、音声認識部6は、その関数値（いわゆるスコア）が最大である音響モデル（単語、または無音（ノイズ））を認識結果として出力する。

【0028】即ち、図4は、図1の音声認識部6の詳細な構成例を示している。

【0029】特徴抽出部5から入力される特徴ベクトル $y$ は、識別関数演算部21-1乃至21- $k$ 、および識別関数演算部21- $s$ に供給される。識別関数演算部21- $k$ （ $k=1, 2, \dots, K$ ）は、 $K$ 個の音響モデルのうちの $k$ 番目に対応する単語を識別するための識別関数 $G_k()$ を記憶しており、特徴抽出部5からの特徴ベクトル $y$ を引数として、識別関数 $G_k(y)$ を演算する。識別関数演算部21- $s$ は、無音モデルに対応する無音区間を識別するための識別関数 $G_s()$ を記憶しており、特徴抽出部5からの特徴ベクトル $y$ を引数として、識別関数 $G_s(y)$ を演算する。

【0030】なお、音声認識部6では、例えば、HMM（Hidden Markov Model）法を用いて、クラスとしての単語または無音の識別（認識）が行われる。

【0031】ここで、図5は、HMMを示している。

【0032】同図において、HMMは、H個の状態 $q_1$ 乃至 $q_H$ を有しており、状態の遷移は、自身への遷移と、右隣の状態への遷移のみが許されている。また、初期状態は、最も左の状態 $q_1$ とされ、最終状態は、最も右の状態 $q_H$ とされており、最終状態 $q_H$ からの状態遷移は禁止されている。このように、自身よりも左にある状態への遷移のないモデルは、left-to-rightモデルと呼ばれ、音声認識では、一般に、left-to-rightモデルが用いられる。

【0033】いま、HMMのkクラスを識別するためのモデル（音響モデル）を、kクラスモデルというすると、kクラスモデルは、例えば、最初に状態 $q_1$ にいる確率（初期状態確率） $\pi_k(q_1)$ 、ある時刻（フレーム）tにおいて、状態 $q_1$ にいて、次の時刻t+1において、状態 $q_1$ に状態遷移する確率（遷移確率） $a_k(q_1, q_1)$ 、および状態 $q_1$ から状態遷移が生じるときに、その状態 $q_1$ が、特徴ベクトルOを出力する確率

$$G_k(y) = \max_{q_1, q_2, \dots, q_T} \pi_k(q_1) \cdot b_k(q_1)(y_1) \cdot a_k(q_1, q_2) \cdot b_k(q_2)(y_2)$$

\*（出力確率） $b_k(q_1)(O)$ によって規定される（ $h = 1, 2, \dots, H$ ）。

【0034】そして、ある特徴ベクトル系列 $O_1, O_2, \dots$ が与えられた場合、例えば、そのような特徴ベクトル系列が観測される確率（観測確率）が最も高いモデルのクラスが、その特徴ベクトル系列の認識結果とされる。

【0035】ここでは、この観測確率が、識別関数 $G_k(y)$ によって求められる。すなわち、識別関数 $G_k(y)$ は、特徴ベクトル（系列） $y = \{y_1, y_2, \dots, y_T\}$ に対する最適状態系列（最適な状態の遷移のしていき方）において、そのような特徴ベクトル（系列） $y = \{y_1, y_2, \dots, y_T\}$ が観測される確率を求めるものとして、次式（1）で与えられる。

【0036】

【数1】

$$\dots \cdot a_k(q_{T-1}, q_T) \cdot b_k(q_T)(y_T) \dots (1)$$

【0037】ここで、連続HMMにおいては、状態 $q_1$ における出力確率 $b_k(q_1)(y_1)$ は、確率分布によって表される。即ち、特徴ベクトル空間上のコンポーネントどうしに相関がないものとして、出力確率 $b_k(q_1)(y_1)$ を規定する確率分布に、正規分布関数※

※を用いることとすると、その正規分布関数 $P(q_1)(d)(y(t)(d))$ は、次式で表すことができる。

【0038】

【数2】

$$= \frac{1}{\sqrt{2\pi(\sum_k(q_1)(d, d))}} e^{-\frac{(\mu_k(q_1)(d) - y(t)(d))^2}{2(\sum_k(q_1)(d, d))}}$$

$$k=1, 2, \dots, K; t=1, 2, \dots, T$$

・・・(2)

但し、式（2）において、 $\mu_k(q_1)(d)$ は、正規分布を規定する平均ベクトルのd番目のコンポーネントを表し、 $\sum_k(q_1)(d, d)$ は、正規分布を規定する分散マトリクスの第d行第d列のコンポーネントを表す。また、 $y(t)(d)$ は、特徴ベクトル $y(t)$ のd番目のコンポーネントを表す。

【0039】kクラスモデルの状態 $q_1$ における出力確率は、式（2）の平均ベクトル $\mu_k(q_1)(d)$ と、分散マトリクス $\sum_k(q_1)(d, d)$ によって規定される。

【0040】なお、HMMは、上述したように、初期状態確率 $\pi_k(q_1)$ 、遷移確率 $a_k(q_1, q_1)$ 、および出力確率 $b_k(q_1)(O)$ によって規定されるが、これらは、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求められる。

【0041】また、HMMとして、図5に示したものを

いる場合には、常に、最も左の状態 $q_1$ から遷移が始まるので、状態 $q_1$ に対応する初期状態確率が1とされ、他の状態に対応する初期状態確率はすべて0とされる。

【0042】さらに、HMMの学習方法としては、例えば、Baum-Welchの再推定法などが知られている。

【0043】図4において、識別関数演算部21-k（ $k=1, 2, \dots, K$ ）は、kクラスモデルについて、あらかじめ学習により求められている初期状態確率 $\pi_k(q_1)$ 、遷移確率 $a_k(q_1, q_1)$ 、および出力確率 $b_k(q_1)(O)$ によって規定される式（2）の識別関数 $G_k(y)$ を記憶しており、特徴抽出部2からの特徴ベクトル $y$ を引数として、識別関数 $G_k(y)$ を演算し、その関数値（上述した観測確率） $G_k(y)$ を、決定部22に出力する。識別関数演算部21-sは、音声モデルとしての、初期状態確率 $\pi_s(q_1)$ 、遷移確率 $a$

40

50

、 $(q_i, q_j)$ 、および出力確率 $b_i(q_i)$  (0) によって規定される式(2)の識別関数 $G_i(y)$ と同様の識別関数 $G_s(y)$ を記憶しており、特徴抽出部2からの特徴ベクトル $y$ を引数として、識別関数 $G_s(y)$ を演算し、その関数値(上述した観測確率) $G_s(y)$ を、決定部22に出力する。

【0044】決定部22では、識別関数演算部21-1\*  

$$C(y) = C_k, \text{ if } G_k(y) = \max_i [G_i(y)]$$

但し、 $C(y)$ は、特徴ベクトル $y$ が属するクラスを識別する識別操作(処理)を行う関数を表す。また、式(3)の第2式の右辺における $\max$ は、それに続く関数値 $G_i(y)$ (ただし、ここでは、 $i = s, 1, 2, \dots, K$ )の最大値を表す。

【0046】決定部22は、式(3)にしたがって、クラスを決定すると、対応する単語(または無音である旨)を、入力された音声の認識結果として出力する。

【0047】図1に戻り、無音モデル補正部7は、ノイズ観測区間抽出部3から入力されるノイズ観測区間 $T_n$ における音声データとしての環境ノイズに基づいて、音声認識部6に記憶されている無音モデルに対応する識別関数 $G_s(y)$ を生成し、この識別関数 $G_s(y)$ によって、音声認識部6に記憶されている無音モデルの適応を行う。

【0048】具体的には、無音モデル補正部7は、ノイ※

$$\mu_{sil} = \frac{1}{M} \sum_{t=1}^M y(t)$$

$$\Sigma_{sil} = \frac{1}{M} \sum_{t=1}^M (y(t) - \mu_{sil})(y(t) - \mu_{sil})^T \quad \dots (4)$$

なお、式(4)における $T$ は、転置を表す。

【0051】そして、無音モデル補正部7は、平均値 $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ で規定される正規分布としての無音モデル $G_s(y)$ によって、識別関数演算部21-sの無音モデル $G_s(y)$ としての識別関数を更新(補正)する。

【0052】次に、図6のフローチャートを参照して、図1の音声認識装置による音声認識処理について説明する。

【0053】フレーム化部2には、マイク1で集音された音声データが入力され、そこでは、音声データがフレーム化され、各フレームの音声データは、観測ベクトル $a$ として、ノイズ観測区間抽出部3、および特徴抽出部5に順次供給される。ノイズ観測区間抽出部3は、ステップS1において、フレーム化部2からの各フレームの音声データをバッファリングする。

【0054】ここで、ノイズ観測区間抽出部3は、少なくともMフレーム以上の音声データを記憶することのできる、図示せぬバッファを有しており、そのバッファの

\*乃至21-k、および識別関数演算部21-sそれぞれからの関数値 $G_i(y)$ (ここでは、関数値 $G_s(y)$ を含むものとする)に対して、例えば、次式(3)に示す決定規則を用いて、特徴ベクトル $y$ 、すなわち、入力された音声に属するクラス(音響モデル)が識別される。

【0045】

【数3】

... (3)

10※ズ観測区間抽出部3から入力されるノイズ観測区間 $T_n$ の音声データ(環境ノイズ)のM個のフレームの各フレームについて、特徴ベクトル $y$ の系列を観測し、その特徴ベクトル $y$ の系列に対して統計的処理を施すことによって、無音モデルを規定する確率分布(無音モデルとしてのHMMの出力確率を規定する確率分布)を生成する。

【0049】即ち、例えば、いま、無音モデルを規定する確率分布が正規分布で表されるとすると、無音モデル補正部7は、ノイズ観測区間 $T_n$ のMフレームの特徴ベクトル $y(t)$ の系列を用い、式(4)にしたがった計算を行うことにより、無音モデル $G_s(y)$ としての正規分布を規定する平均値 $\mu_{sil}$ と、分散マトリクス $\Sigma_{sil}$ を求める。

【0050】

【数4】

記憶容量分の音声データを記憶した後は、最も古い音声データに上書きする形で、新たな音声データを記憶するようになっている。従って、ノイズ観測区間抽出部3では、常に、最新のMフレーム以上の音声データが記憶される。

【0055】その後、ステップS2において、音声区間が開始されたかどうか、即ち、ユーザによって、発話スイッチ4が操作されたかどうかが判定される。ステップS2において、音声区間が開始されていないと判定された場合、ステップS1に戻り、以下、同様の処理を繰り返す。

【0056】また、ステップS2において、音声区間が開始されたと判定された場合、ステップS3に進み、無音モデル補正部7において、無音モデル適応処理が行われる。

【0057】即ち、ステップS2では、ノイズ観測区間抽出部3は、発話スイッチ4がオンとされたタイミング $t_0$ の直前の区間であるノイズ観測区間 $T_n$ の音声データ(環境ノイズ)を、その内蔵するバッファから抽出

し、無音モデル補正部7に供給する。

【0058】無音モデル補正部7は、ノイズ観測区間 $T_n$ の各フレームの音声データの特徴ベクトル $y(t)$ を求め、その特徴ベクトル $(y)$ を用いて、式(4)により、平均値 $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ を求める。そして、無音モデル補正部7は、その平均値 $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ で規定される正規分布 $N(\mu_{sil}, \Sigma_{sil})$ によって、音声認識部6の無音モデル $G_s(y)$ を更新する。

【0059】一方、特徴抽出部5は、発話スイッチ4がオンとされ、音声区間が開始されると、フレーム化部2からの観測ベクトル $a$ としての音声データを音響分析し、その特徴ベクトル $y$ を求め、音声認識部6に供給する。音声認識部6は、ステップS4において、特徴抽出部5からの特徴ベクトル $y$ を用いて、無音と所定数 $K$ の単語それぞれに対応する音響モデルの識別関数の値を演算し、ステップS5に進む。ステップS5では、音声認識部6は、ステップS5で演算した識別関数の関数値が最大となる音響モデルを選択し、対応する単語（または無音）を、音声の認識結果として出力する。

【0060】その後、ステップS6に進み、音声認識処理を終了するかどうか判定され、終了しないと判定された場合、ステップS1に戻り、次の発話について、以下、同様の処理が行われる。

【0061】また、ステップS6において、音声認識処理を終了すると判定された場合、即ち、例えば、ユーザが、音声認識装置の電源をオフする操作を行った場合、処理を終了する。

【0062】次に、上述の場合においては、図7に示すように、各発話の音声区間ごとに、その直前のノイズ観測区間 $T_n$ の音声データ（環境ノイズ）のみに基づいて、無音モデルの適応が行われる。即ち、いま、音声認識装置において音声認識処理が開始されてから、ユーザが行った発話を、第1発話、第2発話、・・・とカウントすることとし、第 $N$ 発話を、現在の音声認識の対象となっている音声の音声区間（注目音声区間）の発話であるとする、第 $N$ 発話の音声区間である注目音声区間の音声の認識には、その注目音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズだけに基づいて生成された無音モデルが用いられる。

【0063】ここで、図7において（後述する図8、図10、図12においても同様）、 $I_n$ は、第 $n$ 発話を表し、 $G_s[I_n]$ は、第 $n$ 発話の音声区間の音声の認識に用いられる無音モデルを表す。

【0064】注目音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズだけに基づいて、無音モデルを生成する場合、前述したように、例えば、注目音声区間の直前にいて、ユーザが、音声の入力に用いるマイクを叩く等し

たときや、観衆が拍手を行ったとき等の、いわば突発的なノイズが生じたときには、その突発的なノイズに基づいて、無音モデルの適応が行われる。

【0065】しかしながら、注目音声区間の、ユーザが発話を行う区間においては、突発的なノイズが存在しなくなるから、突発的なノイズに基づいて生成された無音モデルを用いて、注目音声区間の音声を認識したのでは、認識率が劣化することがある。

【0066】また、比較的定常的なノイズが長時間連続している場合には、注目音声区間のみならず、過去に音声認識の対象とされた音声の音声区間の直前の区間における環境ノイズにも基づいて、無音モデルを生成した方が、環境ノイズをよりの確に表す無音モデルを得ることができると予想され、さらに、そのような無音モデルを用いて、注目音声区間の音声認識を行うことにより、精度の高い音声認識を行うことが可能となる。

【0067】そこで、音声認識装置では、注目音声区間の直前の環境ノイズだけでなく、過去の1以上の音声区間の直前の環境ノイズにも基づいて、以下のような第1乃至第3の3つの適応方法のいずれかにより、注目音声区間の音声を認識するのに用いる無音モデルの適応を行うことが可能となっている。

【0068】即ち、第1の適応方法では、図8に示すように、第 $N$ 発話の音声区間である注目音声区間の直前の環境ノイズと、過去の第1乃至第 $N-1$ 発話の音声区間の直前の環境ノイズから、注目音声区間の音声の認識に用いる無音モデル $G_s[I_n]$ が生成される。

【0069】この場合、図6のステップS3における無音モデル適応処理は、図9のフローチャートに示すように行われる。

【0070】即ち、この場合、ステップS11において、無音モデル補正部7は、第1乃至第 $N$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の音声データ（環境ノイズ）の特徴ベクトル $y(t)$ を計算する。従って、この場合、ノイズ観測区間抽出部3では、注目音声区間である第 $N$ 発話の音声区間の直前の環境ノイズだけでなく、過去の第1乃至第 $N$ 発話の音声区間の直前の環境ノイズも記憶しておく必要がある。

【0071】さらに、無音モデル補正部7は、第1乃至第 $N$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の音声データ（環境ノイズ）の特徴ベクトル $y(t)$ の集合の平均ベクトル $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ を、次式にしたがって計算し、その平均ベクトル $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデル $G_s[I_n]$ とする。

【0072】

【数5】



$$\mu_{sil} = \frac{1}{\sum_{i=1}^N M(i)} \left[ \sum_{i=1}^N W_i \cdot \sum_{t=1}^{M(i)} y(t) [I_i] \right]$$

$$\Sigma_{sil} = \frac{1}{\sum_{i=1}^N M(i)} \left\{ \sum_{i=1}^N \sum_{t=1}^{M(i)} \left[ W_i \cdot y(t) [I_i] - \mu_{sil} \right] \left[ W_i \cdot y(t) [I_i] - \mu_{sil} \right]^T \right\}$$

・・・(5)

【0073】なお、 $M(i)$ は、第 $i$ 発話の音声区間の直前のノイズ観測区間 $T_n$ のフレーム数を表し、本実施の形態では、上述したことから、すべて $M$ フレームである。但し、ノイズ観測区間 $T_n$ のフレーム数は、各発話ごとに、異なるフレーム数とすることが可能である。

【0074】また、 $w_i$ は、第 $i$ 発話の音声区間の直前の環境ノイズに対する重みを表す。この重み $w_i$ は、式 \*

$$\sum_{i=1}^N w_i = 1$$

【0076】さらに、重み $w_i$ は、注目音声区間である第 $N$ 発話の音声区間から離れた音声区間の直前の環境ノイズに対するものほど、小さな値にするようにすることが可能である。

【0077】また、式(5)において、 $y(t)$

$[I_i]$ は、第 $i$ 発話の音声区間の直前の環境ノイズの第 $t$ フレーム(ノイズ観測区間 $T_n$ の第 $t$ フレーム)の特徴ベクトルを表す。

【0078】次に、第2の適応方法では、図10に示すように、過去の第1乃至第 $N-1$ 発話の音声区間の直前の環境ノイズに基づいて、第1の無音モデル $G_{s-1}$ が生成されるとともに、第 $N$ 発話の音声区間である注目音声区間の直前の環境ノイズに基づいて、第2の無音モデル $G_{s-2}$ が生成され、その第1の無音モデル $G_{s-1}$ と、第2の無音モデル $G_{s-2}$ とに基づいて、注目音声区間の音声の認識に用いる無音モデル $G_s$ 、 $[I_N]$ が生成される。

【0079】この場合、図6のステップS3における無音モデル適応処理は、図11のフローチャートに示すように行われる。

【0080】即ち、この場合、ステップS21において、無音モデル補正部7は、第1乃至第 $N-1$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズの特徴ベクトル $y(t)$ を計算する。さらに、無音モデル補正部7は、第1乃至第 $N-1$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズの特徴ベクトル $y(t)$ の集合の平均ベクトル $\mu_{sil-1}$ と分散マトリクス $\Sigma_{sil-1}$ を、式(5)における場合と同様に計算し、その平均ベクトル $\mu_{sil-1}$ と分散マトリクス $\Sigma_{sil-1}$ によって規定される正規分布を、第1の無音モデル $G_{s-1}$ とする。

$$\mu_{sil} = a_{\mu 1} \mu_{sil-1} + b_{\mu 2} \mu_{sil-2}$$

$$\Sigma_{sil} = a_{\Sigma 1} \Sigma_{sil-1} + b_{\Sigma 2} \Sigma_{sil-2}$$

・・・(7)

【0085】ここで、式(7)における $a_{\mu 1}$ 、 $b_{\mu 2}$ 、 $a_{\Sigma 1}$ 、 $b_{\Sigma 2}$ は、いずれも、0以上1以下の範囲の値を

\* (6)を満たすもので、例えば、第 $N$ 発話の音声区間(注目音声区間)の直前の環境ノイズに対する重み $w_N$ は、0.5とし、第1乃至第 $N-1$ 発話の音声区間の直前の環境ノイズに対する重み $w_1$ 乃至 $w_{N-1}$ は、いずれも、0.5/( $N-1$ )とすることが可能である。

【0075】

【数6】

・・・(6)

※【0081】そして、ステップS22に進み、無音モデル補正部7は、注目フレームである第 $N$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズの特徴ベクトル $y(t)$ を計算する。さらに、無音モデル補正部7は、第 $N$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズの特徴ベクトル $y(t)$ の集合の平均ベクトル $\mu_{sil-2}$ と分散マトリクス $\Sigma_{sil-2}$ を、上述の式(4)にしたがって計算し、その平均ベクトル $\mu_{sil-2}$ と分散マトリクス $\Sigma_{sil-2}$ によって規定される正規分布を、第2の無音モデル $G_{s-2}$ とする。

【0082】以上のようにして、第1の無音モデル $G_{s-1}$ と、第2の無音モデル $G_{s-2}$ を得た後は、ステップS23に進み、無音モデル補正部7は、第1の無音モデル $G_{s-1}$ と、第2の無音モデル $G_{s-2}$ とを統合することにより、注目音声区間の音声の認識に用いる無音モデル $G_s$ 、 $[I_N]$ を生成する。

【0083】即ち、無音モデル補正部7は、例えば、式(7)にしたがい、第1の無音モデル $G_{s-1}$ を規定する平均ベクトル $\mu_{sil-1}$ と、第2の無音モデル $G_{s-2}$ を規定する平均ベクトル $\mu_{sil-2}$ とを統合し、平均ベクトル $\mu_{sil}$ を求めるとともに、第1の無音モデル $G_{s-1}$ を規定する分散マトリクス $\Sigma_{sil-1}$ と、第2の無音モデル $G_{s-2}$ を規定する分散マトリクス $\Sigma_{sil-2}$ とを統合し、分散マトリクス $\Sigma_{sil}$ を求める。そして、無音モデル補正部7は、その平均ベクトル $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデル $G_s$ 、 $[I_N]$ とする。

【0084】

【数7】

とる重みであり、式  $a_{\mu 1} + b_{\mu 2} = 1$  と、式  $a_{\Sigma 1} + b_{\Sigma 2} = 1$  を満たすものである。

【0086】環境ノイズが、比較的定常的なものである場合には、重み  $a_{\mu 1}$ 、 $b_{\mu 2}$ 、 $a_{\Sigma 1}$ 、 $b_{\Sigma 2}$  としては、例えば、同一の値を使用することができる。また、環境ノイズが、時間の経過に伴って、比較的变化する場合には、重み  $a_{\mu 1}$ 、 $b_{\mu 2}$ 、 $a_{\Sigma 1}$ 、 $b_{\Sigma 2}$  としては、例えば、 $a_{\mu 1}$  と  $a_{\Sigma 1}$  については、小さな値を、 $b_{\mu 2}$  と  $b_{\Sigma 2}$  については、大きな値を、それぞれ採用することができる。さらに、注目音声区間の直前の環境ノイズが、突発的なものである場合には、重み  $a_{\mu 1}$ 、 $b_{\mu 2}$ 、 $a_{\Sigma 1}$ 、 $b_{\Sigma 2}$  としては、例えば、 $a_{\mu 1}$  と  $a_{\Sigma 1}$  については、大きな値を、 $b_{\mu 2}$  と  $b_{\Sigma 2}$  については、小さな値を、それぞれ採用することができる。

【0087】なお、第1および第2の適応方法においては、過去の音声区間については、注目音声区間より過去の音声区間すべての直前の環境ノイズを用いる他、そのうちの一部の音声区間の直前の環境ノイズを用いて、注目音声区間の音声認識に用いる無音モデルの適応を行うようにすることが可能である。

【0088】次に、第3の適応方法では、図12に示すように、第N発話の音声区間である注目音声区間の直前の環境ノイズに基づいて、無音モデルが生成され、その無音モデルと、過去の音声区間、即ち、図12の実施の形態では、注目音声区間の直前の音声区間（第N-1発話の音声区間）の音声認識に用いられた無音モデルとに基づいて、注目音声区間の音声の認識に用いる無音モデル  $G_s$ 、 $[I_N]$  が生成される。

【0089】この場合、図6のステップS3における無音モデル適応処理は、図13のフローチャートに示すように行われる。

【0090】即ち、この場合、ステップS31において、無音モデル補正部7は、直前の発話、つまり第N-1発話の音声区間の音声認識に用いられた無音モデル  $G_s$ 、 $[I_{N-1}]$  を、音声認識部6（図4）から取得し、ステップS32に進む。

【0091】ステップS32では、無音モデル補正部7は、注目フレームである第N発話の音声区間の直前のノイズ観測区間  $T_n$  の環境ノイズの特徴ベクトル  $y(t)$  を計算する。さらに、無音モデル補正部7は、第N発話の音声区間の直前のノイズ観測区間  $T_n$  の環境ノイズの特徴ベクトル  $y(t)$  の集合の平均ベクトルと分散マトリクスを、上述の式（4）にしたがって計算し、その平均ベクトルと分散マトリクスによって規定される正規分布としての無音モデル  $G_s$ 、 $[I_N]$  を生成する。

【0092】そして、ステップS33に進み、無音モデル補正部7は、第N-1発話の音声区間の音声認識に用いられた無音モデル  $G_s$ 、 $[I_{N-1}]$  と、第N発話の音声区間の直前の環境ノイズだけから得られた無音モデル  $G_s$ 、 $[I_N]$  とを統合することにより、注目音声区間の

音声の認識に用いる無音モデル  $G_s$ 、 $[I_N]$  を生成する。

【0093】即ち、例えば、第N-1発話の音声区間の音声認識に用いられた無音モデル  $G_s$ 、 $[I_{N-1}]$  としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれ  $\mu_{sil-1}$  と  $\Sigma_{sil-1}$  とするとともに、第N発話の音声区間の直前の環境ノイズだけから得られた無音モデル  $G_s$ 、 $[I_N]$  としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれ  $\mu_{sil-2}$  と  $\Sigma_{sil-2}$  とすると、無音モデル補正部7は、ステップS33において、例えば、上述の式（7）にしたがい、平均ベクトル  $\mu_{sil-1}$  と  $\mu_{sil-2}$  とを統合し、平均ベクトル  $\mu_{sil}$  を求めるとともに、分散マトリクス  $\Sigma_{sil-1}$  と  $\Sigma_{sil-2}$  とを統合し、分散マトリクス  $\Sigma_{sil}$  を求める。そして、無音モデル補正部7は、その平均ベクトル  $\mu_{sil}$  と分散マトリクス  $\Sigma_{sil}$  によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデル  $G_s$ 、 $[I_N]$  とする。

【0094】なお、第3の適応方法においては、第N-1発話の音声認識に用いられた無音モデルの他、過去の他の発話の音声認識に用いられた無音モデルを用いて、注目音声区間の音声認識に用いる無音モデルを生成することが可能である。

【0095】以上のように、注目音声区間の直前の環境ノイズだけでなく、過去の1以上の音声区間の直前の環境ノイズにも基づいて、注目音声区間の音声の認識するのに用いる無音モデルの適応を行うようにしたので、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことができ、これにより、無音部分に起因する音声認識性能の劣化を防止（低減）することができる。

【0096】ところで、ノイズ環境下において音声の認識する場合の特徴量（特徴ベクトル）の抽出方法の1つに、例えば、スペクトルサブトラクション（Spectral Subtraction）と呼ばれるものがある。

【0097】スペクトルサブトラクションでは、音声の発話がされる前の入力（音声区間の前の入力）を、ノイズとして、そのノイズの平均スペクトルが算出される。そして、音声区間の音声から、ノイズの平均スペクトルが差し引かれ（Subtract）、その残りを、真の音声成分として、特徴ベクトルが算出される。

【0098】一方、図1の音声認識装置における特徴抽出部5では、各フレームの音声データとしての観測ベクトル  $a$  から、特徴ベクトルが求められるが、このことは、観測ベクトル空間上の、ある点を表す観測ベクトル  $a$  を、特徴ベクトル空間上に写像することにより、その特徴ベクトル空間上の、対応する点を表す特徴ベクトルに変換する処理が行われると考えることができる。

【0099】従って、特徴ベクトルは、特徴ベクトル空間上の、ある1点（観測ベクトル  $a$  に対応する点）を表す。

【0100】スペクトルサブトラクションでは、観測ベクトル  $a$  から、ノイズの平均スペクトル成分が取り除か

10

20

30

40

50

れて、特徴ベクトルが算出されるが、この特徴ベクトルは、上述したように、特徴ベクトル空間上の1点であるため、ノイズの平均的な性質を考慮したものとはなっているが、ノイズの分散などの不規則な性質を考慮したものとはなっていない。

【0101】このため、スペクトルサブトラクション処理後に得られる特徴ベクトルは、観測ベクトル $a$ の特徴を十分に（あるいは、正確に）表現しているとはいえず、そのような特徴ベクトルでは、認識性能を十分に向上させることができないことがある。

【0102】そこで、図14は、本発明を適用した音声認識装置の他の一実施の形態の構成例を示している。なお、図中、図1における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。

【0103】即ち、図14の実施の形態では、図1の特徴抽出部5、音声認識部6、または無音モデル補正部7に替えて、特徴抽出部31、音声認識部32、または無音モデル補正部33がそれぞれ設けられており、さらに、ノイズ観測区間抽出部3が出力する環境ノイズが、無音モデル補正部33だけでなく、特徴抽出部31にも供給されるようになっている。

【0104】但し、ノイズ観測区間抽出部3は、フレーム化部2から入力されるフレーム単位の音声データを、図1における場合よりも長い時間（例えば、2Mフレーム分以上など）だけバッファリングすることができるようになっている。

【0105】即ち、図14の実施の形態においては、ノイズ観測区間抽出部3は、例えば、図15に示すように、発話スイッチ4がオンとされたタイミング $t_0$ からMフレーム分だけ以前のタイミング $t_0$ までを、ノイズ観測区間 $T_n$ とするとともに、さらに、そのノイズ観測区間 $T_n$ からMフレーム分だけ以前のタイミング $t_0$ までをノイズ観測区間 $T_m$ として、その連続する2つのノイズ観測区間 $T_n$ と $T_m$ における2Mフレーム分の観測ベクトル $a$ を抽出して、特徴抽出部31、および無音モデル補正部33に出力する。

【0106】なお、2つのノイズ観測区間 $T_n$ と $T_m$ は、連続していなくてもかまわない。また、ノイズ観測区間 $T_n$ は、上述したように、無音モデルの適応を行うための環境ノイズを得るための区間であるが、ノイズ観測区間 $T_m$ は、後述する特徴分布を抽出するための環境ノイズを得るための区間である。さらに、ここでは、 $2 * x(t) = y(t) - u(t)$

【0113】ただし、ここでは、環境ノイズが不規則な特性を有し、また、観測ベクトル $a(t)$ としての音声データは、真の音声成分に環境ノイズを加算したものであると仮定している。

【0114】一方、ノイズ観測区間抽出部3から入力される音声データとしてのノイズ観測区間 $T_m$ における環

\*つのノイズ観測区間 $T_m$ と $T_n$ を、いずれも、Mフレームで構成するようにしたが、ノイズ観測区間 $T_m$ と $T_n$ のフレーム数は、同一である必要はない。

【0107】特徴抽出部31は、ノイズ観測区間抽出部3から入力されるノイズ観測区間 $T_m$ と $T_n$ のうちの前半のノイズ観測区間 $T_m$ の環境ノイズだけが存在する音声データに基づいて、フレーム化部2から入力される、タイミング $t_0$ 以降の音声区間の観測ベクトル $a$ から環境ノイズ成分を除去して、その特徴量を抽出する。

【0108】即ち、特徴抽出部31は、例えば、図1の特徴抽出部5と同様に、観測ベクトル $a$ としての音声データをフーリエ変換し、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトル $y$ を算出する。さらに、特徴抽出部31は、観測ベクトル $a$ としての音声データに含まれる真の音声成分を、その特徴量の空間（特徴ベクトル空間）に写像したときに得られる、その特徴ベクトル空間上の分布を表すパラメータ（以下、特徴分布パラメータと記述する） $Z$ を、特徴ベクトル $y$ とノイズ観測区間 $T_m$ の環境ノイズに基づいて算出し、音声認識部32に供給する。

【0109】即ち、図16は、図14の特徴抽出部31の詳細な構成例を示している。なお、図中、図3の特徴抽出部5における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、特徴抽出部31は、特徴分布パラメータ算出部42とノイズ特性算出部43が新たに設けられている他は、図3の特徴抽出部5と同様に構成されている。

【0110】フレーム化部2から入力される観測ベクトル $a$ は、特徴抽出部31において、パワースペクトラム分析部11に供給され、特徴ベクトル $y$ としてのパワースペクトラムとされる。なお、ここでは、1フレームの音声データとしての観測ベクトル $a$ が、 $D$ 個のコンポーネントからなる特徴ベクトル（ $D$ 次元の特徴ベクトル）に変換されるものとする。

【0111】ここで、第 $t$ フレームの観測ベクトル $a(t)$ から得られる特徴ベクトル $y(t)$ のうち、真の音声のスペクトル成分を $x(t)$ と、環境ノイズのスペクトル成分を $u(t)$ と表す。この場合、真の音声のスペクトル成分 $x(t)$ は、次式(8)で表される。

【0112】

【数8】

$$\dots (8)$$

境ノイズは、特徴抽出部31において、ノイズ特性算出部43に供給される。ノイズ特性算出部43では、ノイズ観測区間 $T_m$ における環境ノイズの特性が求められる。

【0115】即ち、ここでは、音声区間における環境ノイズのパワースペクトラム $u(t)$ の分布が、その音声

区間の直前のノイズ観測区間  $T_m$  における環境ノイズと同一であり、かつ、その分布が正規分布であると仮定して、ノイズ特性算出部 43 において、その正規分布を規定する、環境ノイズの平均ベクトル  $\mu'$  と分散マトリクス\*

$$\mu'(i) = \frac{1}{M} \sum_{t=1}^M y(t)(i)$$

$$\Sigma'(i, j) = \frac{1}{M} \sum_{t=1}^M (y(t)(i) - \mu'(i))(y(t)(j) - \mu'(j))$$

... (9)

【0117】ただし、 $\mu'(i)$  は、平均ベクトル  $\mu'$  の  $i$  番目のコンポーネントを表す ( $i = 1, 2, \dots, D$ )。また、 $y(t)(i)$  は、第  $t$  フレームの特徴ベクトルの  $i$  番目のコンポーネントを表す。さらに、 $\Sigma'(i, j)$  は、分散マトリクス  $\Sigma'$  の、第  $i$  行、第  $j$  列のコンポーネントを表す ( $j = 1, 2, \dots, D$ )。

$$\Sigma'(i, j) = 0, i \neq j$$

\*  $\Sigma'$  が、次式 (9) にしたがって求められる。

【0116】

【数9】

※【0118】ここで、計算量の低減のために、環境ノイズについては、特徴ベクトル  $y$  の各コンポーネントが、互いに無相関であると仮定する。この場合、次式に示すように、分散マトリクス  $\Sigma'$  は、対角成分以外は 0 となる。

【0119】

【数10】

... (10)

【0120】なお、環境ノイズについて、特徴ベクトル  $y$  の各コンポーネントが、互いに無相関であると仮定しなくても、計算量は増加するが、以下説明する処理を行うことは可能である。

【0121】ノイズ特性算出部 43 は、以上のようにして、環境ノイズの特性としての、正規分布を規定する平均ベクトル  $\mu'$  および分散マトリクス  $\Sigma'$  を求め、特徴分布パラメータ算出部 42 に供給する。

【0122】一方、パワースペクトラム分析部 11 の出力、すなわち、環境ノイズを含む音声区間の音声の特徴ベクトル  $y$  も、特徴分布パラメータ算出部 42 に供給される。特徴分布パラメータ算出部 42 は、パワースペクトラム分析部 11 からの特徴ベクトル  $y$ 、およびノイズ★

$$\xi(t)(i) = E[x(t)(i)]$$

$$= E[y(t)(i) - u(t)(i)]$$

$$= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i)) \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i)$$

$$= \frac{y(t)(i) \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) - \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)}$$

$$= y(t)(i) - \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)}$$

... (11)

【0125】

【数12】

20★特性算出部 43 からの環境ノイズの特性（ここでは、環境ノイズを表す正規分布を規定する平均ベクトル  $\mu'$  と分散マトリクス  $\Sigma'$ ）に基づいて、真の音声のパワースペクトラムの分布（推定値の分布）を表す特徴分布パラメータを算出する。

【0123】即ち、特徴分布パラメータ算出部 42 は、真の音声のパワースペクトラムの分布が正規分布であるとして、その平均ベクトル  $\xi$  と分散マトリクス  $\Psi$  を、特徴分布パラメータとして、次式 (11) 乃至 (14) にしたがって計算する。

【0124】

【数11】

30

<sup>21</sup>  
i=jのとき

22

$$\begin{aligned}\Psi(t)(i, j) &= V[x(t)(i)] \\ &= E[(x(t)(i))^2] - (E[x(t)(i)])^2 \\ &= (E[(x(t)(i))^2] - (\xi(t)(i))^2)\end{aligned}$$

i≠jのとき

$$\Psi(t)(i, j) = 0 \quad \dots (12)$$

【0126】

\*10\* 【数13】

$$\begin{aligned}E[(x(t)(i))^2] &= E[(y(t)(i) - u(t)(i))^2] \\ &= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i))^2 \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\ &= \frac{1}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \times \left\{ (y(t)(i))^2 \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) \right. \\ &\quad - 2y(t)(i) \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i) \\ &\quad \left. + \int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i) \right\} \\ &= (y(t)(i))^2 - 2y(t)(i) \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\ &\quad + \frac{\int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \quad \dots (13)\end{aligned}$$

【0127】

※ ※ 【数14】

$$P(u(t)(i)) = \frac{1}{\sqrt{2\pi \Sigma'(i, i)}} e^{-\frac{1}{2\Sigma'(i, i)}(u(t)(i) - \mu'(i))^2} \quad \dots (14)$$

【0128】ここで、 $\xi(t)(i)$ は、第tフレームにおける平均ベクトル $\xi(t)$ のi番目のコンポーネントを表す。また、 $E[\ ]$ は、 $[\ ]$ 内の平均値を意味する。 $x(t)(i)$ は、第tフレームにおける真の音声のパワースペクトラム $x(t)$ のi番目のコンポーネントを表す。さらに、 $u(t)(i)$ は、第tフレームにおける環境ノイズのパワースペクトラムのi番目のコンポーネントを表し、 $P(u(t)(i))$ は、第tフレームにおける環境ノイズのパワースペクトラムのi番目のコンポーネントが $u(t)(i)$ である確率を表す。ここでは、環境ノイズの分布として正規分布を仮定しているので、 $P(u(t)(i))$ は、式(14)に示したように表される。

【0129】また、 $\Psi(t)(i, j)$ は、第tフレームにおける分散マトリクス $\Psi(t)$ の、第i行、第j列のコンポーネントを表す。さらに、 $V[\ ]$ は、 $[\ ]$ 内の分散を表す。

【0130】特徴分布パラメータ算出部42は、以上のようにして、各フレームごとに、平均ベクトル $\xi$ および分散マトリクス $\Psi$ を、真の音声の特徴ベクトル空間上での分布（ここでは、真の音声の特徴ベクトル空間上での分布が正規分布であると仮定した場合の、その分布）を表す特徴分布パラメータとして求める。

【0131】特徴分布パラメータ算出部42は、音声区間の各フレームについて求めた特徴分布パラメータを、音声認識部32に出力する。すなわち、いま、音声区間

40

50

がTフレームであったとし、そのTフレームそれぞれにおいて求められた特徴分布パラメータを、 $z(t) = \{\xi(t), \Psi(t)\}$  ( $t=1, 2, \dots, T$ )と表すと、特徴分布パラメータ算出部42は、特徴分布パラメータ(系列) $Z = \{z(1), z(2), \dots, z(T)\}$ を、音声認識部32に供給する。

【0132】図14に戻り、音声認識部32は、特徴抽出部31から入力される特徴分布パラメータZを、所定数Kの音響モデルと1個の無音モデルのうちのいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。

【0133】即ち、音声認識部32は、例えば、無音区間に対応する識別関数と、所定数Kの単語それぞれに対応する識別関数とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部31からの特徴分布パラメータZを引数として計算する。そして、その関数値が最大である音響モデル(単語、または無音(ノイズ))が認識結果として出力される。

【0134】ここで、図17は、図14の音声認識部32の詳細な構成例を示している。なお、図中、図4の音声認識部6における場合と対応する部分については、同一の符号を付してある。即ち、音声認識部32は、基本\*

$$G_k(Z) = \max_{q_1, q_2, \dots, q_T} \pi_k(q_1) \cdot b'_k(q_1)(z_1) \cdot a_k(q_1, q_2) \cdot b'_k(q_2)(z_2)$$

$$\dots \cdot a_k(q_{T-1}, q_T) \cdot b'_k(q_T)(z_T) \dots (15)$$

【0138】ここで、 $b'_k(q_i)(z_j)$ は、出力が $z_j$ で表される分布であるときの出力確率を表す。式(1)で説明したように、状態遷移時に各特徴ベクトルを出力する確率である出力確率 $b_k(S)(O_i)$ に( $S$ は状態を表す)、特徴ベクトル空間上のコンポーネントに相関がないものとして、正規分布関数を用いることとした場合、入力 $z_i$ で表される分布であるときは、出力確率 $b'_k(S)(z_i)$ は、平均ベクトル $\mu_k(S)$  ※

$$b'_k(S)(z_t) = \int P^f(t)(x) P_k^m(S)(x) dx \\ = \prod_{i=1}^D P(S)(i)(\xi(t)(i), \Psi(t)(i, i))$$

$$k=1, 2, \dots, K; S=q_1, q_2, \dots, q_T; t=1, 2, \dots, T$$

... (16)

【0140】ただし、式(16)における積分の積分区間は、D次元の特徴ベクトル空間(ここでは、パワースペクトラム空間)の全体である。

【0141】また、式(16)において、 $P(S)$

※的に、図4の音声認識部6と同様に構成されている。

【0135】但し、識別関数演算部21-1乃至21-k、および識別関数演算部21-sには、特徴抽出部31の特徴分布パラメータ算出部42が出力する特徴分布パラメータZが供給されるようになっており、識別関数演算部21-k ( $k=1, 2, \dots, K, s$ )は、特徴分布パラメータZを引数とする識別関数 $G_k(Z)$ を、音響モデルとして記憶している。

【0136】図17の実施の形態において、音声認識部32が、例えば、図4の音声認識部6と同様に、HMM法を用いて、クラスとしての単語または無音の識別(認識)を行う場合、音声認識部32は、音響モデルとしてのHMMにおいて、特徴分布パラメータの系列 $Z = \{z_1, z_2, \dots, z_T\}$ が観測される観測確率を、識別関数 $G_k(Z)$ によって求める。即ち、この場合、識別関数 $G_k(Z)$ は、特徴分布パラメータの系列 $Z = \{z_1, z_2, \dots, z_T\}$ に対する最適状態系列において、そのような特徴分布パラメータ(の系列) $Z = \{z_1, z_2, \dots, z_T\}$ が観測される確率を求めるものとして、次式(15)で与えられる。

【0137】

【数15】

$$\dots \cdot a_k(q_{T-1}, q_T) \cdot b'_k(q_T)(z_T) \dots$$

... (15)

※と分散マトリクス $\Sigma_k(S)$ とによって規定される確率密度関数 $P_k(S)(x)$ 、および第tフレームの特徴ベクトル(ここでは、パワースペクトラム)xの分布を表す確率密度関数 $P^f(t)(x)$ を用いて、次式(16)により求めることができる。

【0139】

【数16】

(i) ( $\xi(t)(i), \Psi(t)(i, i)$ )は、次式(17)で表される。

【0142】

【数17】

$$P(S)(i)(\xi(t)(i), \Psi(t)(i, i))$$

$$= \frac{1}{\sqrt{2\pi(\sum_k(S)(i, i) + \Psi(t)(i, i))}} e^{-\frac{(\mu_k(S)(i) - \xi(t)(i))^2}{2(\sum_k(S)(i, i) + \Psi(t)(i, i))}} \dots (17)$$

【0143】ただし、 $\mu_k(S)(i)$ は、平均ベクトル $\mu_k(S)$ の*i*番目のコンポーネントを、 $\Sigma_k(S)(i, i)$ は、分散マトリクス $\Sigma_k(S)$ の、第*i*行第*i*列のコンポーネントを、それぞれ表す。そして、*k*クラ

【0144】なお、HMMは、上述した場合と同様に、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求めておく。

【0145】ここで、特徴分布パラメータ*Z*に基づく音声認識に用いられる出力確率を規定する式(17)の確率分布は、特徴分布パラメータ*Z*の分散 $\Psi(t)(i, i)$ を0とすると、特徴ベクトルの分散を考慮しない場合の連続HMMにおける出力確率を規定する式(2)の確率分布に一致する。

【0146】決定部22は、図4における場合と同様に、識別関数演算部21-1乃至21-*k*、および識別関数演算部21-*s*それぞれからの関数値 $G_k(Z)$

(関数値 $G_k(Z)$ を含む)に対して、上述の式(3)と同様の決定規則を用いて、特徴分布パラメータ*Z*、即ち、入力された音声に属するクラス(音響モデル)を識\*  
 $[F_1(y), F_2(y), \dots, F_M(y)]$

【0150】ここで、環境ノイズの特徴分布パラメータ $F_1(y)$ は、ユーザの音声のない部分、つまり無音(正確には、環境ノイズが存在する)の特徴ベクトルの分布を表すから、以下、適宜、無音特徴分布とも記述する。

$$G_s(y) = V(F_1(y), F_2(y), \dots, F_M(y))$$

【0153】但し、*V*は無音特徴分布 $\{F_1(Z), i=1, 2, \dots, M\}$ を無音モデル $G_s(Z)$ に写像する補正関数(写像関数)である。

【0154】この写像は、無音特徴分布の記述によって★

$$G_s(y) = \sum_{i=1}^M \beta_i(F_1(y), F_2(y), \dots, F_M(y), M) \cdot F_i(y) \\ = \sum_{i=1}^M \beta_i \cdot F_i(y) \dots (20)$$

【0156】但し、 $\beta_i(F_1(y), F_2(y), \dots, F_M(y), M)$ は、ノイズ観測区間 $T_n$ の第*i*フレームから得られる無音特徴分布 $F_i(y)$ に対する重み関数であり、以下、 $\beta_i$ と記述する。なお、重み関数

\*別し、音声認識結果として出力する。

【0147】図14に戻り、無音モデル補正部33は、ノイズ観測区間抽出部3から入力されるノイズ観測区間 $T_m$ と $T_n$ における音声データとしての環境ノイズに基づいて、音声認識部32に記憶されている無音モデルに対応する識別関数 $G_s(Z)$ を生成し、この識別関数 $G_s(Z)$ によって、音声認識部32に記憶されている無音モデルの適応を行う。

【0148】具体的には、無音モデル補正部33では、ノイズ観測区間抽出部3から入力される後半のノイズ観測区間 $T_n$ の音声データ(環境ノイズ)の*M*個のフレームの各フレームについて、特徴ベクトル $y$ が観測され、さらに、特徴抽出部31における場合と同様に、前半のノイズ観測区間 $T_m$ における環境ノイズを用いて、後半のノイズ観測区間 $T_n$ の各フレーム#*i*における環境ノイズの特徴分布パラメータの、次式で示される系列が生成される。

【0149】

【数18】

... (18)

※【0151】次に、無音モデル補正部33は、無音特徴分布を、次式(19)に従い、無音モデルに対応する確率分布 $G_s(Z)$ に写像する。

【0152】

【数19】

... (19)

★様々な方法が考えられるが、例えば、次式を採用することができ。

【0155】

【数20】

$\beta_i$ は、次式(21)の条件を満足するものである。

【0157】

【数21】

$$\sum_{i=1}^M \beta_i (F_1(y), F_2(y), \dots, F_M(y), M) = \sum_{i=1}^M \beta_i \equiv 1 \quad \dots (21)$$

【0158】ノイズ観測区間 $T_n$ における各フレームの特徴ベクトル $y$ を構成するコンポーネントが無相関であれば、無音特徴分布 $\{F_i(y), i=1, 2, \dots, M\}$ は、平均ベクトル $\mu_i$ と分散マトリクス $\Sigma_i$ で規定される正規分布 $N(\mu_i, \Sigma_i)$ となる。

【0159】この場合、無音モデル補正部33は、ノイズ観測区間 $T_n$ の各フレームから得られる無音特徴分布\*10

$$\begin{aligned} \mu_{sil} &= \frac{a}{M} \sum_{i=1}^M \mu_i \\ \Sigma_{sil} &= \frac{b}{M} \sum_{i=1}^M \Sigma_i \end{aligned} \quad \dots (22)$$

【0161】ここで、係数 $a$ および $b$ としては、例えば、シミュレーションにより最適な値を決定することができる。

【0162】なお、無音特徴分布 $F_i(y)$ から、無音モデル $G_s(Z)$ を生成する方法は、上述の方法に限定されるものではなく、例えば、本件出願人が先に出願した特願2000-276856号（特願平11-375766号を基礎とする国内優先権主張出願）等の開示されている各種の方法を採用することができる。

【0163】ところで、上述のように、無音特徴分布を用いて、無音モデルの適応を行う場合においても、特徴ベクトルを用いて、無音モデルの適応を行う場合と同様に、注目音声区間の直前の環境ノイズだけでなく、過去※

\*  $F_i(y)$ としての正規分布を規定する平均ベクトル $\mu_i$ と分散マトリクス $\Sigma_i$ を用い、例えば、次式にしたがって、無音モデル $G_s(Z)$ を表す正規分布を規定する平均ベクトル $\mu_{sil}$ と、 $\Sigma_{sil}$ を演算する。

【0160】

【数22】

※の1以上の音声区間の直前の環境ノイズにも基づき、上述の第1乃至第3の3つの適応方法（図8乃至図13）によって、注目音声区間の音声認識するのに用いる無音モデルの適応を行うことが可能である。

20 【0164】即ち、例えば、いま、第 $i$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の第 $j$ フレームの環境ノイズから得られる無音特徴分布を、 $F_j[I_i]$ と表すとすると、本実施の形態では、特徴ベクトルが $D$ 次のコンポーネントで構成されるから、無音特徴分布 $F_j[I_i]$ は、次式に示すような $D$ 次のコンポーネントで表される。

【0165】

【数23】

$$F_j[I_i] = \begin{bmatrix} f_1(i, j) \\ \vdots \\ f_d(i, j) \\ \vdots \\ f_D(i, j) \end{bmatrix} \quad \dots (23)$$

【0166】また、第 $i$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の第 $j$ フレームの環境ノイズから得られる無音特徴分布 $F_j[I_i]$ の $d$ 番目のコンポーネント $f_d(i, j)$ は、上述したことから、式(24)に示すように、平均値 $\mu_d(i, j)$ と、分散 $\sigma_d^2(i, j)$ に★

$$f_d(i, j) = N(\mu_d(i, j), \sigma_d^2(i, j)) \quad \dots (24)$$

★よって規定される正規分布 $N(\mu_d(i, j), \sigma_d^2(i, j))$ で表すことができる。

【0167】

【数24】

【0168】この場合、第1の適法方法（図8）では、無音モデル補正部33は、第1乃至第 $N$ 発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズから得られる無音特徴分布の $d$ 番目のコンポーネントの平均値 $\mu_{sil}$  ☆

☆ $(d)$ と分散 $\sigma_{sil}^2(d)$ を、次式にしたがって計算する。

【0169】

【数25】

$$\begin{aligned} \mu_{sil}(d) &= \frac{1}{\sum_{i=1}^N M(i)} \left[ \sum_{i=1}^N W_i \cdot \sum_{j=1}^{M(i)} \mu_d(i, j) \right] \\ \sigma_{sil}^2(d) &= \frac{1}{\sum_{i=1}^N M(i)} \left[ \sum_{i=1}^N W_i \cdot \sum_{j=1}^{M(i)} \sigma_d^2(i, j) \right] \end{aligned} \quad \dots (25)$$



【0170】そして、無音モデル補正部33は、次式で示される平均ベクトル $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ によって規定される正規分布を、注目音声区間の音声の認識\*

$$\mu_{sil} = \begin{bmatrix} \mu_{sil}(1) \\ \vdots \\ \mu_{sil}(d) \\ \vdots \\ \mu_{sil}(D) \end{bmatrix}, \quad \Sigma_{sil} = \begin{bmatrix} \sigma_{sil}^2(1) & & 0 \\ & \ddots & \\ 0 & & \sigma_{sil}^2(D) \end{bmatrix} \quad \dots (26)$$

【0172】次に、第2の適応方法(図10)では、無音モデル補正部33は、第1乃至第N-1発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズから得られる無音特徴分布 $F_j[I_1]$ 乃至 $F_j[I_{N-1}]$ を計算する。さらに、無音モデル補正部33は、無音特徴分布 $F_j[I_1]$ 乃至 $F_j[I_{N-1}]$ の平均ベクトル $\mu_{sil-1}$ と分散マトリクス $\Sigma_{sil-1}$ を、式(25)および(26)に※

$$G_{s-1} = N(\mu_{sil-1}, \Sigma_{sil-1})$$

【0174】また、無音モデル補正部33は、注目フレームである第N発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズから得られる無音特徴分布 $F_j[I_N]$ 20のd番目のコンポーネントについて、そのノイズ観測区★

$$\mu_{sil-2}(d) = \frac{1}{M(N)} \sum_{d=1}^{M(N)} \mu_d(N, j)$$

$$\sigma_{sil-2}^2(d) = \frac{1}{M(N)} \sum_{d=1}^{M(N)} \sigma_d^2(N, j)$$

【0176】さらに、無音モデル補正部33は、式(29)で示される平均ベクトル $\mu_{sil-2}$ と分散マトリクス $\Sigma_{sil-2}$ によって規定される正規分布 $N(\mu_{sil-2}, \Sigma_{sil-2})$ を、式(30)に示すように、第2の無音モデル

$$\mu_{sil-2} = \begin{bmatrix} \mu_{sil-2}(1) \\ \vdots \\ \mu_{sil-2}(d) \\ \vdots \\ \mu_{sil-2}(D) \end{bmatrix}, \quad \Sigma_{sil-2} = \begin{bmatrix} \sigma_{sil-2}^2(1) & & 0 \\ & \ddots & \\ 0 & & \sigma_{sil-2}^2(D) \end{bmatrix} \quad \dots (29)$$

【0178】

$$G_{s-2} = N(\mu_{sil-2}, \Sigma_{sil-2})$$

◆ ◆ 【数30】

... (30)

【0179】以上のようにして、第1の無音モデル $G_{s-1}$ と、第2の無音モデル $G_{s-2}$ を得た後は、無音モデル補正部33は、第1の無音モデル $G_{s-1}$ と、第2の無音モデル $G_{s-2}$ とを統合することにより、注目音声区間の音声の認識に用いる無音モデル $G_s[I_N]$ を生成する。

【0180】即ち、無音モデル補正部33は、例えば、上述の式(7)にしたがい、第1の無音モデル $G_{s-1}$ を規定する平均ベクトル $\mu_{sil-1}$ と、第2の無音モデル $G_{s-2}$ を規定する平均ベクトル $\mu_{sil-2}$ とを統合し、平均ベクトル $\mu_{sil}$ を求めるとともに、第1の無音モデル $G_{s-1}$ ※

$$G_{sil} = N(\mu_{sil}, \Sigma_{sil})$$

※に用いる無音モデル $G_s[I_N]$ とする。

【0171】

【数26】

※における場合と同様に計算し、その平均ベクトル $\mu_{sil-1}$ と分散マトリクス $\Sigma_{sil-1}$ によって規定される正規分布 $N(\mu_{sil-1}, \Sigma_{sil-1})$ を、式(27)に示すように、第1の無音モデル $G_{s-1}$ とする。

【0173】

【数27】

... (27)

★間 $T_n$ における平均値 $\mu_{sil-2}(d)$ と分散 $\sigma_{sil-2}^2(d)$ を、次式にしたがって計算する。

【0175】

【数28】

... (28)

☆ル $G_{s-2}$ とする。

【0177】

【数29】

☆30

※を規定する分散マトリクス $\Sigma_{sil-1}$ と、第2の無音モデル $G_{s-2}$ を規定する分散マトリクス $\Sigma_{sil-2}$ とを統合し、分散マトリクス $\Sigma_{sil}$ を求める。そして、無音モデル補正部33は、その平均ベクトル $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ によって規定される正規分布 $N(\mu_{sil}, \Sigma_{sil})$ を、式(31)に示すように、注目音声区間の音声の認識に用いる無音モデル $G_s[I_N]$ とする。

【0181】

【数31】

... (31)

【0182】次に、第3の適応方法(図12)では、無音モデル補正部33は、直前の発話、つまり第N-1発話の音声区間の音声認識に用いられた無音モデル $G_s [I_{N-1}]$ を、音声認識部32(図14)から取得する。

【0183】さらに、無音モデル補正部33は、注目フレームである第N発話の音声区間の直前のノイズ観測区間 $T_n$ の環境ノイズから得られる無音特徴分布 $F_j [I_N]$ の、そのノイズ観測区間 $T_n$ における平均ベクトルと分散マトリクスを、上述の式(28)および(29)にしたがって計算し、その平均ベクトルと分散マトリクスによって規定される正規分布としての無音モデル $G_s' [I_N]$ を生成する。

【0184】そして、無音モデル補正部33は、第N-1発話の音声区間の音声認識に用いられた無音モデル $G_s [I_{N-1}]$ と、第N発話の音声区間の直前の環境ノイズだけから得られた無音モデル $G_s' [I_N]$ とを統合することにより、注目音声区間の音声の認識に用いる無音モデル $G_s [I_N]$ を生成する。

【0185】即ち、例えば、第N-1発話の音声区間の音声認識に用いられた無音モデル $G_s [I_{N-1}]$ としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれ $\mu_{sil-1}$ と $\Sigma_{sil-1}$ とするとともに、第N発話の音声区間の直前の環境ノイズだけから得られた無音モデル $G_s' [I_N]$ としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれ $\mu_{sil-2}$ と $\Sigma_{sil-2}$ とすると、無音モデル補正部33は、例えば、上述の式

(7)にしたがい、平均ベクトル $\mu_{sil-1}$ と $\mu_{sil-2}$ とを統合し、平均ベクトル $\mu_{sil}$ を求めるとともに、分散マトリクス $\Sigma_{sil-1}$ と $\Sigma_{sil-2}$ とを統合し、分散マトリクス $\Sigma_{sil}$ を求める。そして、無音モデル補正部33は、その平均ベクトル $\mu_{sil}$ と分散マトリクス $\Sigma_{sil}$ によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデル $G_s [I_N]$ とする。

【0186】以上のように、特徴分布パラメータを用いる場合においても、注目音声区間の直前の環境ノイズだけでなく、過去の1以上の音声区間の直前の環境ノイズにも基づいて、注目音声区間の音声の認識するのに用いる無音モデルの適応を行うようにすることで、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことができ、これにより、無音部分に起因する音声認識性能の劣化を防止(低減)することができる。

【0187】なお、第i発話の音声区間の直前の環境ノイズから得られる無音特徴分布 $F_j [I_N]$ の、そのノイズ観測区間 $T_n$ における平均ベクトルと分散マトリクスを、上述の式(28)および(29)にしたがって計算する場合においては、そのノイズ観測区間 $T_n$ におけるMフレームそれぞれから得られる時系列の無音特徴分布 $F_1 [I_1], F_2 [I_1], \dots, F_M [I_1]$ は、平等に扱うのではなく、例えば、対応する音声区間に近いも

のほど重みをおいて扱うことが可能である。このような重み付けの方法の詳細については、例えば、上述の特願2000-276856号等に記載されている。

【0188】次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0189】そこで、図18は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0190】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0191】あるいはまた、プログラムは、フロッピー(登録商標)ディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体111に、一時的あるいは永続的に格納(記録)しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。

【0192】なお、プログラムは、上述したようなリムーバブル記録媒体111からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部108で受信し、内蔵するハードディスク105にインストールすることができる。

【0193】コンピュータは、CPU(Central Processing Unit)102を内蔵している。CPU102には、バス101を介して、入出力インタフェース110が接続されており、CPU102は、入出力インタフェース110を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部107が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM(Random Access Memory)104にロードして実行する。これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に

じて、例えば、入出力インタフェース110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0194】ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0195】また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0196】なお、本実施の形態では、音響モデルとして、HMMを採用し、HMM法に基づいて音声認識を行うようにしたが、音響モデル、および音声認識のアルゴリズムは、特に限定されるものではない。

【0197】また、本発明は、特徴ベクトルと、特徴分布パラメータの両方を用いて音声認識を行う場合にも適用可能である。

【0198】

【発明の効果】本発明のモデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置によれば、音声区間の直前の区間で観測される音声データが抽出され、抽出データとして出力される。そして、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応が行われる。従って、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことが可能となる。

【図面の簡単な説明】

【図1】本発明を適用した音声認識装置の第1実施の形態の構成例を示すブロック図である。

【図2】図1のノイズ観測区間抽出部3の処理を説明する図である。

【図3】特徴抽出部5の構成例を示すブロック図である。

【図4】音声認識部6の構成例を示すブロック図であ

る。

【図5】HMMを示す図である。

【図6】音声認識処理を説明するフローチャートである。

【図7】各発話に対して、無音モデルの適応を行う様子を示す図である。

【図8】第1の無音モデルの適応方法を説明する図である。

【図9】第1の無音モデルの適応方法による無音モデル適応処理を説明するフローチャートである。

【図10】第2の無音モデルの適応方法を説明する図である。

【図11】第2の無音モデルの適応方法による無音モデル適応処理を説明するフローチャートである。

【図12】第3の無音モデルの適応方法を説明する図である。

【図13】第3の無音モデルの適応方法による無音モデル適応処理を説明するフローチャートである。

【図14】本発明を適用した音声認識装置の第2実施の形態の構成例を示すブロック図である。

【図15】図14のノイズ観測区間抽出部3の処理を説明する図である。

【図16】特徴抽出部31の構成例を示すブロック図である。

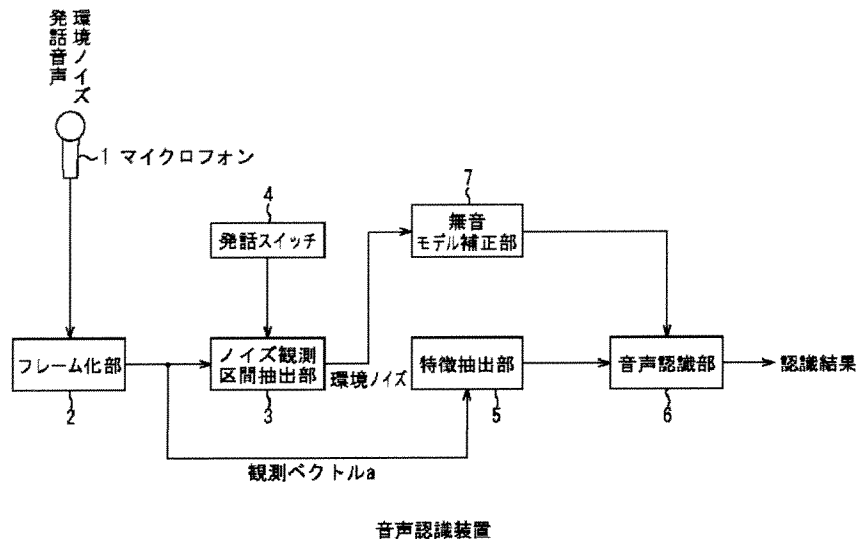
【図17】音声認識部32の構成例を示すブロック図である。

【図18】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

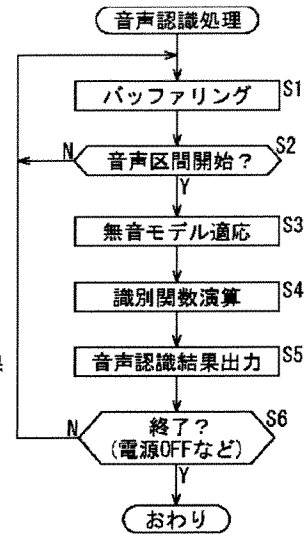
【符号の説明】

1 マイクロフォン、 2 フレーム化部、 3 ノイズ観測区間抽出部、 4 発話スイッチ、 5 特徴抽出部、 6 音声認識部、 7 無音モデル補正部、 11 パワースペクトラム分析部、 21-1乃至21-N、 21-s 識別関数演算部、 22 決定部、 31 特徴抽出部、 32 音声認識部、 33 無音モデル補正部、 42 特徴分布パラメータ算出部、 43 ノイズ特性算出部、 101 バス、 102 CPU、 103 ROM、 104 RAM、 105 ハードディスク、 106 出力部、 107 入力部、 108 通信部、 109 ドライブ、 110 入出力インタフェース、 111 リムーバブル記録媒体

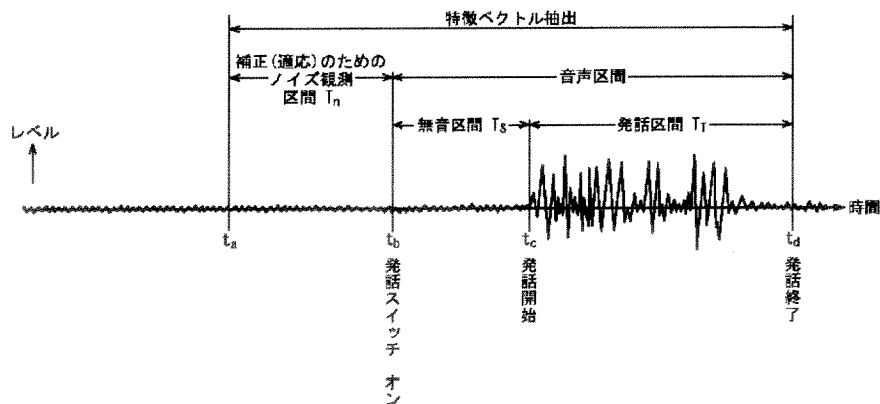
【図1】



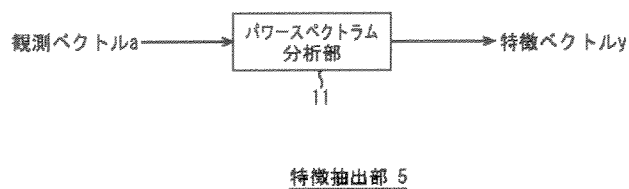
【図6】



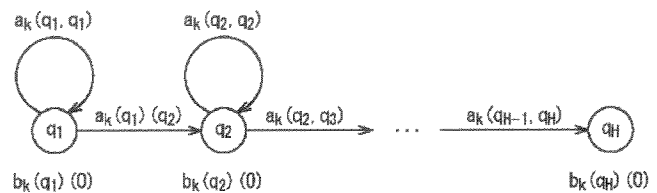
【図2】



【図3】

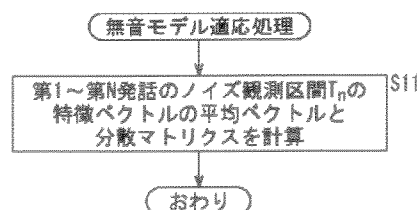


【図5】

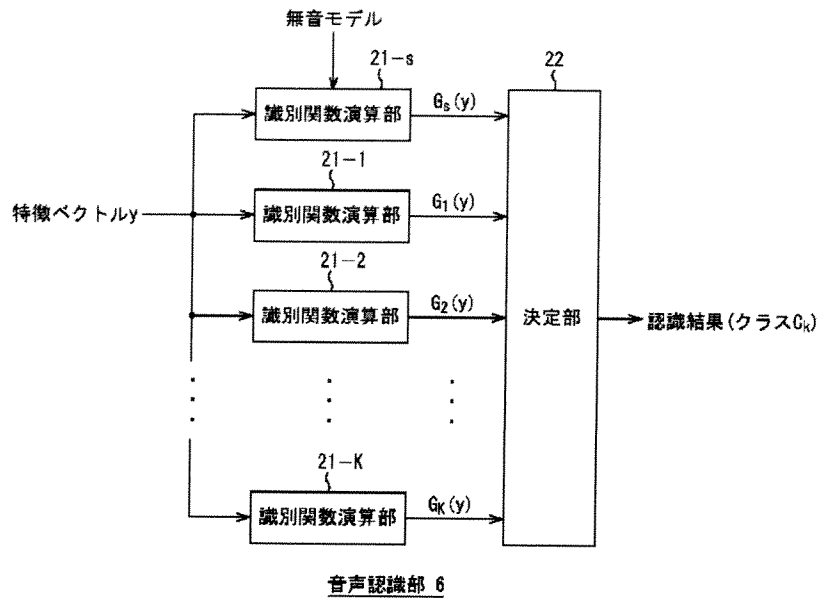


HMM(left-to-right モデル)

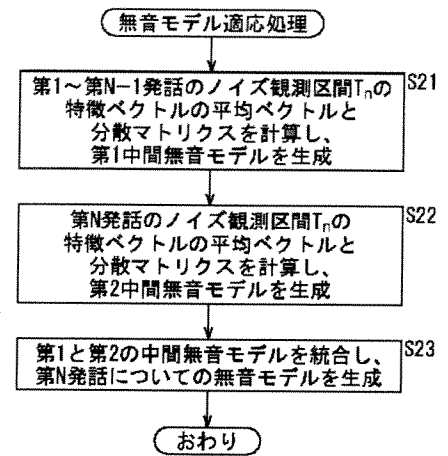
【図9】



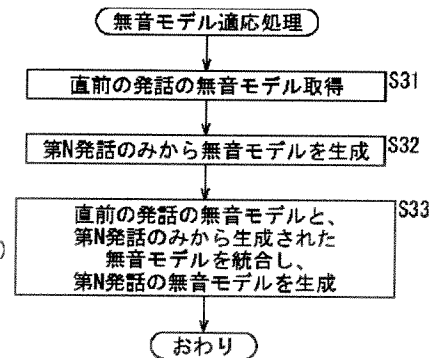
【図4】



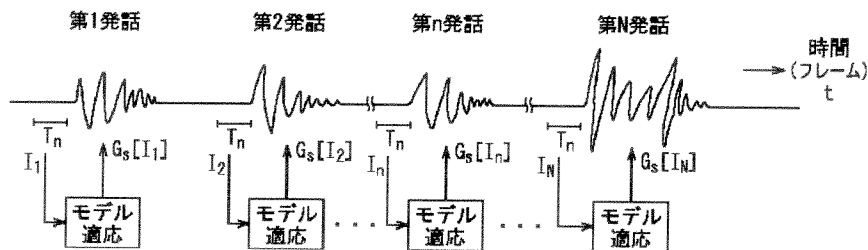
【図11】



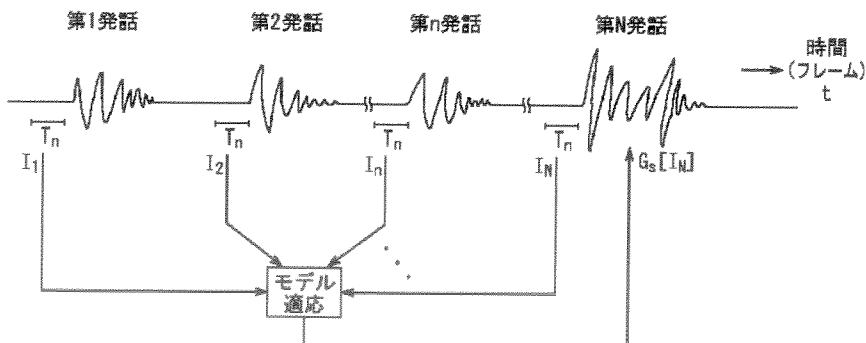
【図13】



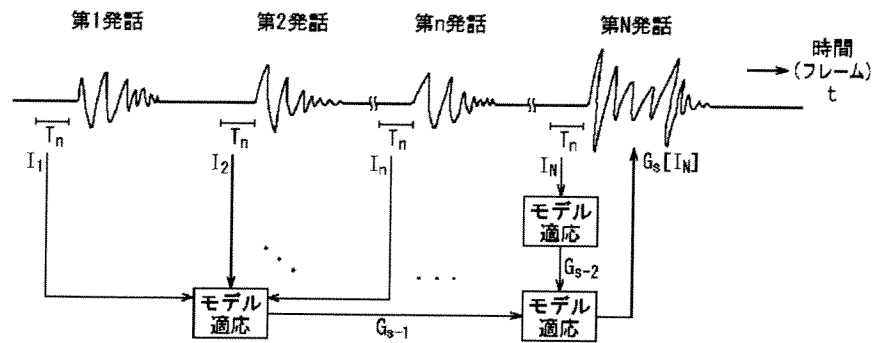
【図7】



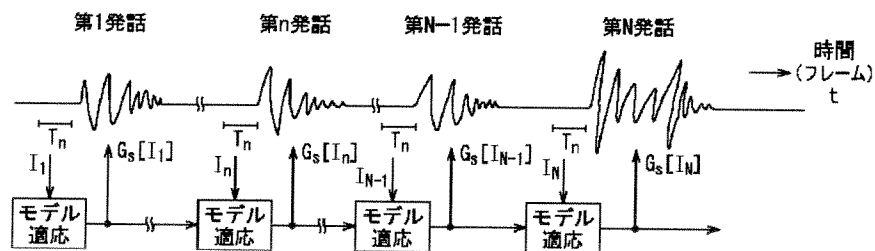
【図8】



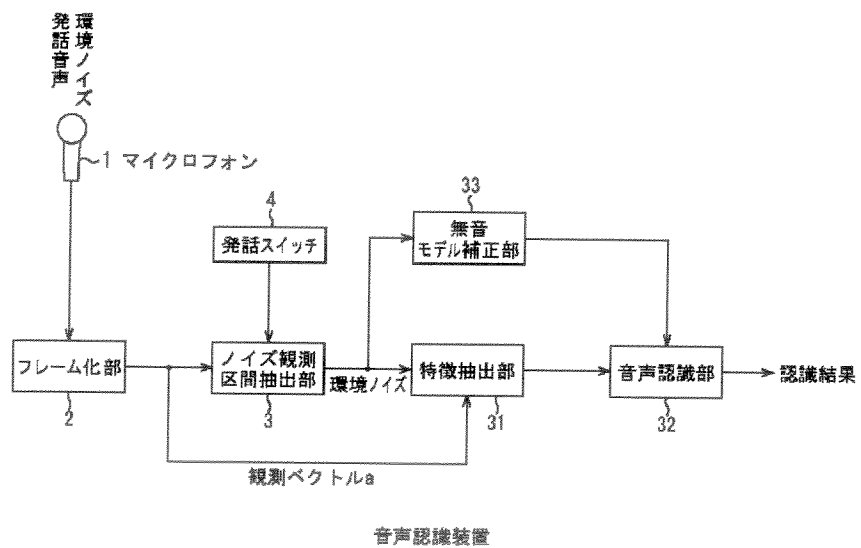
【図10】



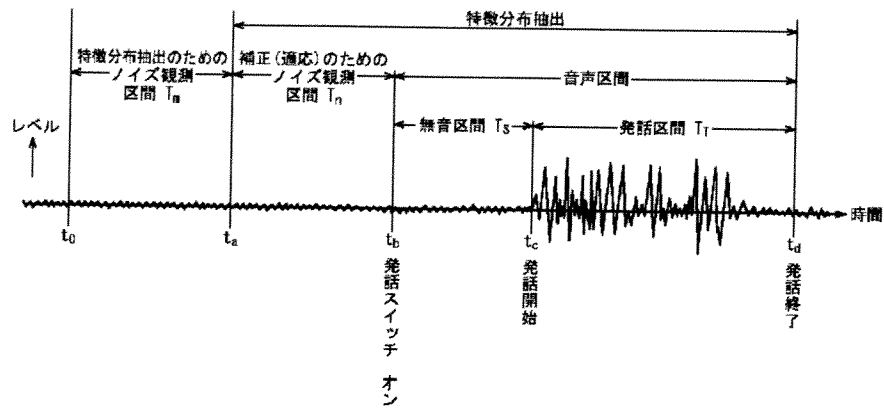
【図12】



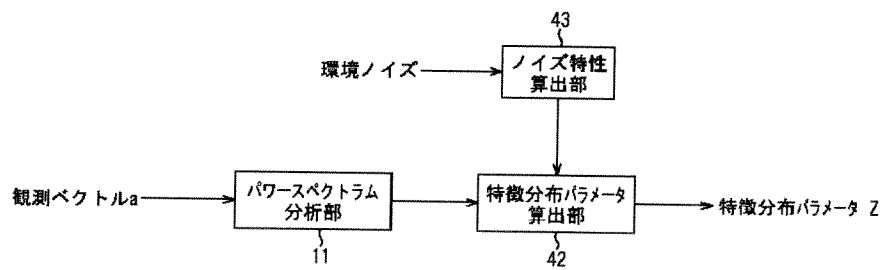
【図14】



【図15】

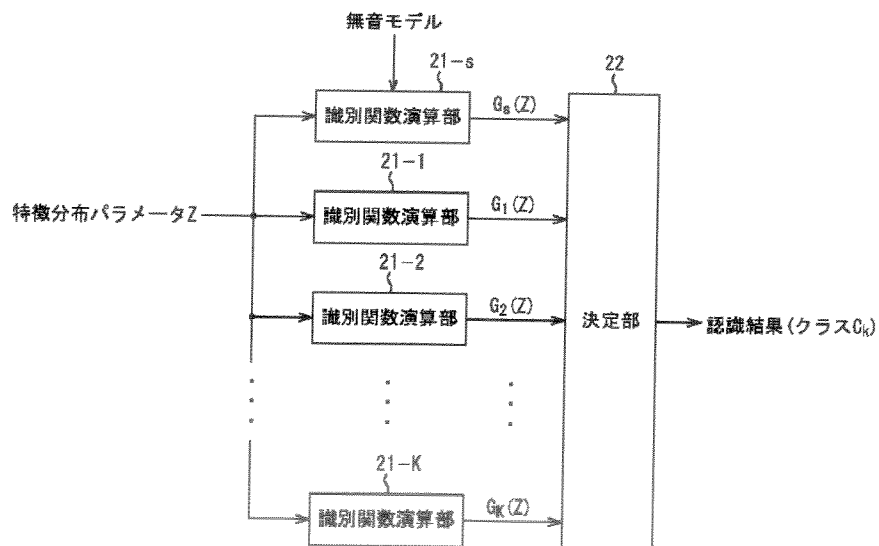


【図16】



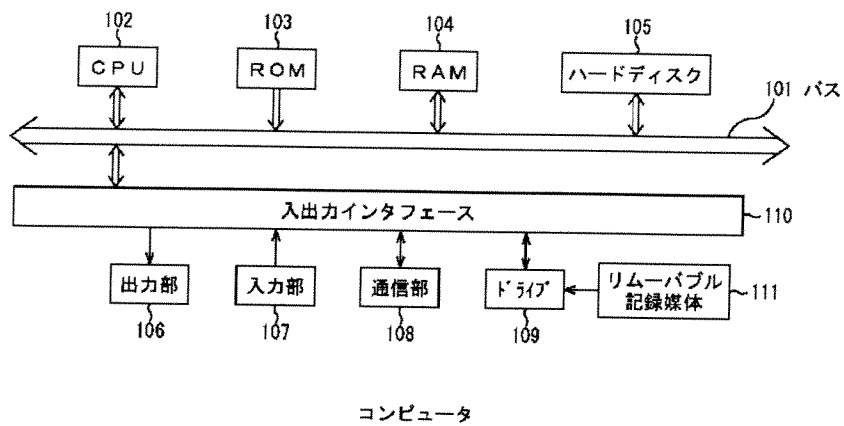
特徴抽出部 31

【図17】



音声認識部 32

【図18】



【手続補正書】

【提出日】平成12年11月28日(2000.11.28)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】発明の名称

【補正方法】変更

【補正内容】

【発明の名称】モデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置